

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Multimodal Ensemble Fusion Deep Learning Using Histopathological Images and Clinical Data For Glioma Subtype Classification

SATOSHI SHIRAE¹, SHYAM SUNDAR DEBSARKAR², HIROHARU KAWANAKA¹, BRUCE ARONOW^{3,4,5}, V. B. SURYA PRASATH^{3,4,5}

¹Graduate School of Engineering, Mie University, Mie 514-8507, Japan (e-mail: 424m220@m.mie-u.ac.jp, kawanaka@elec.mie-u.ac.jp)

²Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221 USA (e-mail: debsars@mail.uc.edu)

³Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229 USA (e-mail: bruce.aronow@cchmc.org, surya.prasath@cchmc.org)

⁴Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45257 USA

⁵Department of Biomedical Informatics, College of Medicine, University of Cincinnati, OH 45267 USA

Corresponding author: V. B. Surya Prasath (e-mail: surya.iit@gmail.com).

ABSTRACT Glioma is the most common malignant tumor of the central nervous system, and diffuse Glioma is classified as grades II-IV by world health organization (WHO). In the the cancer genome atlas (TCGA) Glioma dataset, grade II and III Gliomas are classified as low-grade glioma (LGG), and grade IV Gliomas as glioblastoma multiforme (GBM). In clinical practice, the survival and treatment process with Glioma patients depends on properly diagnosing the subtype. With this background, there has been much research on Glioma over the years. Among these researches, the origin and evolution of whole slide images (WSIs) have led to many attempts to support diagnosis by image analysis. On the other hand, due to the disease complexities of Glioma patients, multimodal analysis using various types of data rather than a single data set has been attracting attention. In our proposed method, multiple deep learning models are used to extract features from histopathology images, and the features of the obtained images are concatenated with those of the clinical data in a fusion approach. Then, we perform patch-level classification by machine learning (ML) using the concatenated features. Based on the performances of the deep learning models, we ensemble feature sets from top three models and perform further classifications. In the experiments with our proposed ensemble fusion AI (EFAI) approach using WSIs and clinical data of diffuse Glioma patients on TCGA dataset, the classification accuracy of the proposed multimodal ensemble fusion method is 0.936 with an area under the curve (AUC) value of 0.967 when tested on a balanced dataset of 240 GBM, 240 LGG patients. On an imbalanced dataset of 141 GBM, 242 LGG patients the proposed method obtained the accuracy of 0.936 and AUC of 0.967. Our proposed ensemble fusion approach significantly outperforms the classification using only histopathology images alone with deep learning models. Therefore, our approach can be used to support the diagnosis of Glioma patients and can lead to better diagnosis.

INDEX TERMS Deep Learning, Machine Learning, CNN, Transformer, Artificial Intelligence, Histopathology, Whole Slide Images, Clinical Data, Glioma, Ensemble, Multimodal, Classification.

I. INTRODUCTION

Glioma is the most common tumor arising in the central nervous system, accounting for approximately 80% of malignant primary tumors of the brain. Diffuse Gliomas, which are the subject of this research, are characterized by invasive growth into the CNS parenchyma and are most common in adult patients. Diffuse Gliomas are classified as grade II-IV by WHO based on the presence of marked mitotic activity, necrosis, and vivid microvascular proliferation. In the cancer

genome atlas (TCGA) [1] diffuse glioma dataset, grade IV gliomas are classified as glioblastoma multiforme (GBM), while grade II and III gliomas are classified as low grade glioma (LGG) [2]–[5].

GBM accounts for approximately 60% of all brain tumors in adults and is a fatal malignancy with a very poor prognosis. The median survival time from diagnosis for patients with GBM is rather short, approximately 14-15 months. GBM is a very rare tumor with a worldwide incidence of less than 10

What specific original CNN Architecture did they use to get better results? How does this relate to this paper?

per 100,000 people. GBM is also a very heterogeneous tumor characterized by multiple molecular markers, hemorrhages, necrosis, and cystic lesions [6]–[10]. Digital pathology plays an important role in modern clinical practice with the origin and advancement of whole slide images (WSIs). WSIs are a complex and rich source of information, containing pathological information at multiple scales (e.g. 4x and 20x) [11]–[13]. However, humans cannot typically extract all of this information from WSIs, and further the human evaluation of images is highly subjective. Therefore, the importance of automatic image analysis by applying artificial intelligence (AI) to WSIs has increased in recent years. Using AI, it is possible to automatically extract information that cannot be identified by humans, thereby reducing the time and cost of patient diagnosis and processing, as well as errors caused by human bias [14], [15].

Among such applications of AI, a great deal of research has been conducted in recent years on the use of deep learning for the analysis of WSIs [16]. However, most of these researches use a single deep learning model, which is not sufficient to completely capture the characteristics of lesion sites from WSIs. Therefore, our proposed method uses multiple deep learning models for feature extraction of WSIs in a fusion framework. This ensembling approach provides better imaging features from the WSIs than compared to individual deep learning models. In addition, we consider that the disease complexity cannot be grasped in detail only from the information provided by images. Therefore, in this research, we apply the analysis of clinical data as well as histopathological images such as WSIs to the classification of diffuse Glioma subtypes (GBM and LGG). Our method uses multiple deep learning models for feature extraction and aims to improve classification performance by using multiple types of data.

Im et al [17] applied a transfer learning method to the classification of Glioma subtypes to resolve the imbalance between positive and negative data, which is very difficult to collect in the clinical field. The authors used ResNet50V2 as the base model for transfer learning. In an experiment to classify whether a patient had Oligodendrogloma, a subtype of Glioma, the classification accuracy was 0.7870 and the F1 score was 0.7231. Slide-level classification by majority vote of patch-level classification for Oligodendrogloma-positive and Oligodendrogloma-negative resulted in a classification accuracy of 0.8727. The authors concluded that these results have the potential to improve the productivity of pathologists and suggest their usefulness in future clinical practice. Yang et al [18] used pathological images and clinical information to predict breast cancer recurrence. The authors used ResNet50 to extract features from pathological images. For clinical information, the authors used a random forest to verify the importance of the features and to determine the data to be used. The final risk of recurrence was evaluated using these features. In an experiment using pathological images and clinical information, the AUC for the test dataset was 0.72, which was much higher than that predicted by the clinical information alone. The authors stated that the results show

new possibilities for clinical applications of deep learning. Yonekura et al [4] implemented a new CNN architecture to classify Glioma subtypes GBM and LGG. The authors originally constructed a convolutional layer, ReLU, and a pooling layer. Experiments using histopathology images of TCGA Glioma patients showed an average classification accuracy of 96.5%. This result is more accurate than several existing CNN architectures. The authors found that the proposed CNN is useful for medical decision analysis of brain tumors. Rathore et al [19] used pathology image features and clinical information to grade Glioma. The features of pathological images include statistical features such as skewness, morphological features such as eccentricity, and textural features such as cluster shade. Morphological features of the pathological images were extracted using the k-means clustering algorithm. SVM was used for final classification. In an experiment using data from TCGA patients with Glioma, the best classification was achieved when all features were used, with a classification accuracy of 91.48%. The authors stated that the proposed method has the potential to serve as a practical pathology image-based biomarker.

Our study introduces a significant advancement in the classification of gliomas through the strategic ensemble of top-performing deep learning models, specifically focusing on two distinct pathways: one for convolutional neural networks (CNNs) [20], [21] and another for Transformers [22]. This methodology focuses on aggregating the strengths of the top-performing CNNs, to enhance their capability in extracting and analyzing visual features from whole slide images (WSIs). Concurrently, another ensemble approach to the top-performing Transformer models was used, capitalizing their exceptional ability to process and interpret sequential data. This dual-pathway strategy is designed to harness the specialized advantages of both CNNs and transformers, optimizing their respective performances in the complex task of glioma subtype classification. We perform three different sets of experiments, wherein we first test the efficacy of various state-of-the-art DL models (CNNs, Transformers), second using global average pooling, and principal component analysis (PCA) to map the top three DL models' features into ensemble classification with various machine learning classifiers. Finally, we use an ensemble fusion AI (EFAI) approach that combines top three DL models along with clinical data fusion. This approach allows for a more focused exploration of the data, potentially leading to improved diagnostic accuracy, reliability, and a deeper understanding of glioma characteristics. Further, we conduct the experiments on balanced and imbalanced datasets to test the robustness of the models.

Rest of our work is organized as follows. Section II provides the details of our ensemble AI model with WSI and clinical data fusion for brain glioma classification along with initial experiments using either data modalities alone. Section III provides detailed experimental results using different DL and ML models as well as the results of our proposed ensemble fusion approach. Section IV concludes the paper.

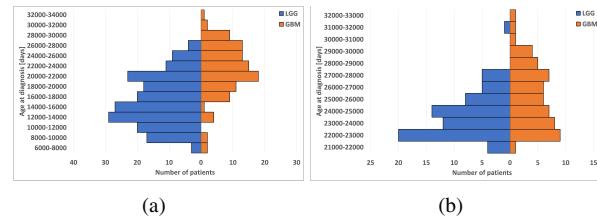


FIGURE 1. Comparison of age in days at diagnosis on (a) balanced (patients:281, slides:480), and (b) imbalanced (patients:125, slides:353).

II. ENSEMBLE FUSION DEEP LEARNING FOR GLIOMA SUBTYPE CLASSIFICATION

A. DATA AND PREPROCESSING

In our work, we utilized Glioblastoma datasets retrieved from the Genomic Data Commons data portal (<https://portal.gdc.cancer.gov/>) and the publicly accessible resources provided by the TCGA database (<https://portal.gdc.cancer.gov>). The datasets comprised primarily of Whole Slide Images (WSIs), accompanied by pertinent clinical data for the patients. WSIs, which are high-resolution images of tissue samples mounted on glass slides, formed the core of our study material. We downloaded a total of 1,839 WSIs, with 298 slides categorized under Glioblastoma Multiforme (GBM) and 1,541 under Low Grade Glioma (LGG), setting the stage for a binary classification task. To accommodate the requirements of our classification task, we curated both balanced and unbalanced datasets. For the balanced dataset, we randomly selected 240 WSIs each from the GBM and LGG categories. In contrast, the unbalanced dataset was composed of 141 GBM and 212 LGG WSIs, again chosen randomly, with a specific emphasis on samples from patients older than 60 years. Given the substantial size of the WSIs, with average dimensions exceeding tens of thousands of pixels in both width and height, adaptation for deep learning models necessitated the division of each WSI into smaller patches of 256×256 pixels. This process also involved the elimination of white spaces and non-tissue areas within the WSIs. Ultimately, this preparation yielded a total of 30,720 patches for the balanced dataset and 22,592 patches for the unbalanced dataset, ready for further analysis. The distribution of age_at_diagnosis for GBM and LGG patients in the balanced and imbalanced datasets used in this research is shown in the Figure 1. In the graphs, the horizontal axis represents the number of patients, and the vertical axis represents the age at diagnosis in days of life.

B. PROPOSED ENSEMBLE FUSION AI (EFAI) APPROACH

The overview of our proposed method is shown in Fig. 2. First, our proposed method extracts features from a histopathological image using multiple deep learning models (Experiment 1 - CNN and Transformer models). In this process, different features can be extracted for a single histopathological image depending on the number of deep learning models used. The purpose of this method is to obtain features that cannot be captured by a single deep learning

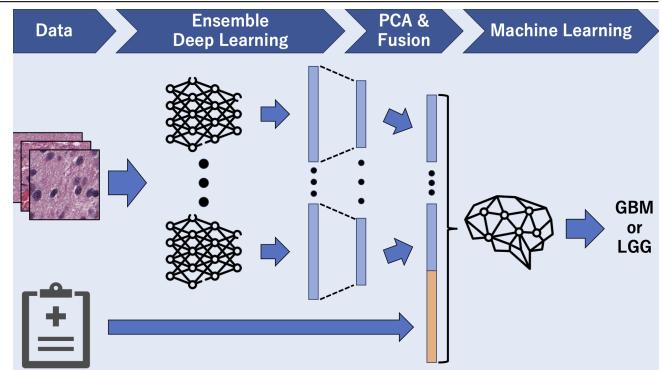


FIGURE 2. Overview of the proposed ensemble fusion AI (EFAI) approach with whole-slide imaging and associated clinical data.

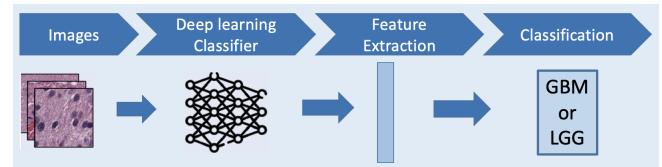


FIGURE 3. Experiment 1 : DL models (CNNs, Transformers) for both balanced and imbalanced for binary classification.

model. Next, feature selection and feature concatenation are performed on the extracted features based on the top 3 models in both balanced and imbalanced datasets. Principal component analysis (PCA) is used to reduce the dimensionality and make the size of the features extracted from each model uniform [23]. This prevents the dimensionality of the features from increasing too much due to the use of multiple deep learning models. After selecting top DL models we tested different machine learning classifiers [24] on the top 3 ensembled CNN and Transformer models' features (Experiment 2 - Ensemble + ML). Finally, the features from the histopathological images are concatenated with the clinical data to complete the patch-by-patch features used for the final classification. The features containing the histopathological images and clinical data are used to perform patch-level classification using different machine learning classifiers (Experiment 3 - Ensemble + Fusion + ML). The proposed ensemble fusion AI (EFAI) approach integrates features extracted from histopathological images and clinical data to improve classification performance. Deep learning models (CNNs and Transformers) extract high-dimensional image features, which are then concatenated and reduced using Principal Component Analysis (PCA) to optimize feature representation. Meanwhile, clinical features such as age, gender, and genetic markers are normalized and combined with the reduced image features to form a comprehensive feature vector. This fused feature set is then processed by machine learning classifiers (e.g., LightGBM, Random Forest) to generate the final classification output. This fusion strategy ensures that clinically relevant patient data refines the decision boundary of the image-based classifier, enhancing diagnostic accuracy.

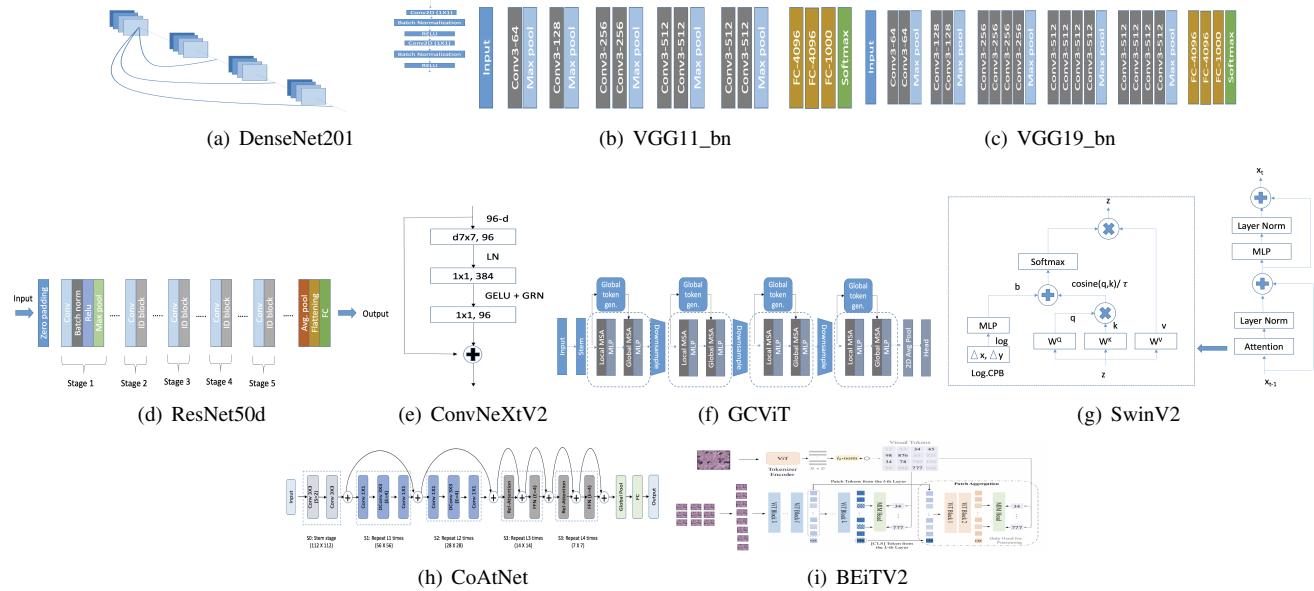


FIGURE 4. We used the following DL models. Top row: Top CNN models - (a) DenseNet201 [25], (b) VGG11_bn [26], (c) VGG19_bn [26], (d) ResNet50d [27], (e) ConvNeXtV2 [28]. Bottom Row: Top Transformers - (f) GCViT [29], (g) SwinV2 [30], (h) CoAtNet [31], (i) BEiT2V [32].

C. DEEP LEARNING MODEL SELECTION (EXPERIMENT 1)

Since our proposed method uses multiple deep learning models for feature extraction of tissue images in an ensemble approach, it is necessary to select optimal models. In recent years, deep learning models (DL) have been evolving at a rapid pace, and there are many CNN-based models [20], [21] and transformer-based models such as Vision Transformer (ViT) [22] that are effective for many vision tasks. Each deep learning model has its own unique structure, and researchers need to select the appropriate model for each individual task. Although deep learning models are frequently used in research on the classification of patient diseases using histopathology images, the process and reasons for selecting the model are often not mentioned. Therefore, we performed classification using histopathological images as input for each single deep learning model and compared the results to determine the top model best suited for our task, see Figure 3.

We compared the classification performance of 18 deep learning models pre-trained by ImageNet. Among the models used for comparison, the CNN-based models are ConvNeXtV2 [28], DenseNet201 [25], EfficientNetV2_b0 [33], EfficientNetV2_b3 [33], MobileNetV3 [34], ResNet50d [27], ResNet101d [27], ResNeXt101 [35], VGG11_bn [26], and VGG19_bn [26]. and the Transformer-based models are BEiT2V [32], CoAtNet [31], CrossViT [36], DEiT [37], GCViT [29], MobileViTV2 [38], Swin TransformerV2 [30] and ViT [22]. Figure 4 shows the basic blocks and networks we utilized here. These models were selected from a range of DL models, some of which are frequently used in pathological image classification research and some of which are not yet widely used.

D. FEATURES OF SELECTED DEEP LEARNING MODEL (EXPERIMENT 2 ENSEMBLE)

The patch images were input to each of the top 3 models (based on maximum accuracy) selected by the classification performance comparison, and feature maps were extracted. Each top model was pre-trained using the patch image dataset. Global average pooling (GAP) converts the extracted feature maps into feature vectors. Since the feature maps extracted from each deep learning model have different sizes, the lengths of the converted feature vectors are also different, as shown in the Table 1. Note that in CNNs the top 3 models for balanced dataset are DenseNet201, VGG19_bn, VGG11_bn, and imbalanced dataset are CNN models are VGG11_bn, ResNet50d, ConvNeXtV2, respectively. In Transformers the top 3 models for balanced dataset are GCViT, SwinV2, CoAtNet and for imbalanced dataset are GCViT, SwinV2, BEiT2V, respectively. If the feature vectors extracted from each model are concatenated with their lengths as they are, the result is a non-uniform vector. In addition, the feature vectors after concatenation become very long, which is expected to be detrimental due to the curse of dimensionality. Therefore, principle component analysis (PCA) was used to reduce the dimensionality and align the lengths of the feature vectors extracted from each model. The procedure from feature extraction to dimensionality reduction of feature vectors was performed for all patch images. Figure 5 shows the overall ensembling approach.

E. MACHINE LEARNING CLASSIFIERS

We conducted experiments with different machine learning classifiers on ensembled image-based deep learning features as well clinical data fusion to achieve the final prediction. We utilized four different classifiers, namely SVM [39],

TABLE 1. Feature length of top 3 CNN and Transformer-based models trained on balanced and imbalanced datasets. Note that in CNNs the top 3 models for balanced dataset are DenseNet201, VGG19_bn, VGG11_bn, and imbalanced dataset are VGG11_bn, ResNet50d, ConvNeXtV2, respectively. In Transformers the top 3 models for balanced dataset are GCViT, SwinV2, CoAtNet and for imbalanced dataset are GCViT, SwinV2, BEiT2, respectively.

CNNs	DenseNet201	VGG19_bn	VGG11_bn	ResNet50d	ConvNeXtV2
Length	1920	512	512	2048	768
Transformers	GCViT	SwinV2	CoAtNet	BEiT2	
Length	512	768	512	768	

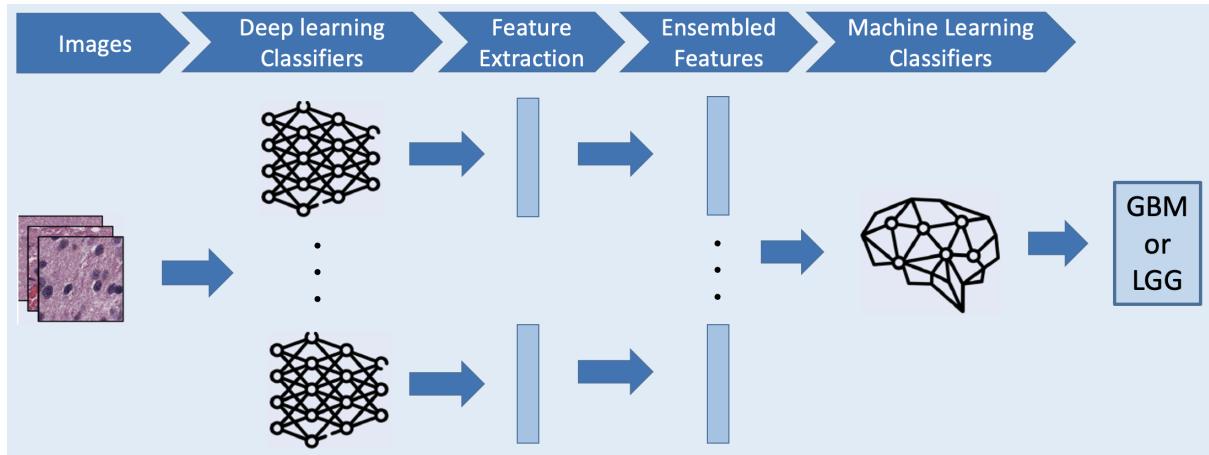


FIGURE 5. Experiment 2 : Ensembling top 3 models (CNNs, Transformers) for both balanced and imbalanced for binary classification.

MLP [40], LightGBM [41], and Random Forest (RF) [42], each selected for their distinct advantages in predictive analytics, see Figure 6. SVM and MLP, while potentially more complex and computationally demanding, are capable of delivering superior performance metrics. In contrast, LightGBM and Random Forest offer speed and efficiency, making them ideal for managing large and intricate datasets. However, MLP and LightGBM require meticulous hyperparameter tuning to prevent overfitting. SVMs are particularly effective in classification scenarios with well-defined margin separation. MLPs are adept at tackling complex pattern recognition tasks. LightGBM is known for its robust performance with categorical and voluminous data, whereas Random Forest provides flexibility across various data types, though it can be resource-intensive.

While this study focuses on glioma subtype classification, the proposed ensemble fusion AI (EFAI) framework is inherently adaptable to other medical classification tasks that utilize histopathology images and clinical data. The methodology of deep learning-based feature extraction from WSIs and fusion with clinical attributes is a general approach that can be applied to various diseases, including breast, lung, and prostate cancer, where histopathological and clinical data play crucial diagnostic roles. Additionally, this multimodal fusion strategy can be extended to predict disease progression, treatment outcomes, or patient prognosis in different medical domains with appropriate dataset availability. Further ensembling approaches [43], [44] such as voting, weighted ranking and hybrid stacking can also be incorporated into our EFAI

along state-of-the-art DL models.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In the evaluation process, 10% of the patched image dataset was reserved as test data to assess the classification performance of each model. The remaining dataset underwent five-fold cross-validation to fine-tune hyperparameters, as illustrated in Fig. 7. During cross-validation, each model was trained and validated over 100 epochs, utilizing RAdam for optimization and ExponentialLR to adjust the initial learning rate at each epoch. To evaluate classification performance, five metrics were used: Accuracy, Recall, Precision, F1-score, and AUC. Classification comparisons primarily focused on the highest Accuracy values achieved. Among CNN-based models evaluated on the balanced dataset, DenseNet201, VGG19_bn, and VGG11_bn emerged as the top performers in descending order. For Transformer-based models, GCViT, Swin TransformerV2, and CoAtNet ranked the highest. Conversely, on the imbalanced dataset, the top CNN-based models were VGG11_bn, ResNet50d, and ConvNeXtV2, while the leading Transformer models were GCViT, Swin Transformer V, and BEiT2, all listed in descending order of performance. For each model category, the iteration with the highest Accuracy was preserved as the feature extractor for subsequent methods. Upcoming experiments will replicate these methodologies using both CNN-based and Transformer-based models, with a comparative analysis of the outcomes.

TABLE 2. Results of various deep learning models with balanced dataset

	Data	Accuracy		Recall		Precision		F1-score		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
ConvNeXtV2	WSIs	0.876	0.890	0.766	0.801	0.906	0.927	0.829	0.852	0.975	0.979
DenseNet201	WSIs	0.894	0.912	0.789	0.822	0.933	0.949	0.855	0.881	0.898	0.921
EfficientNetV2_b0	WSIs	0.884	0.887	0.806	0.814	0.890	0.904	0.846	0.849	0.930	0.937
EfficientNetV2_b3	WSIs	0.875	0.883	0.795	0.809	0.879	0.888	0.835	0.845	0.922	0.926
MobileNetV3	WSIs	0.888	0.899	0.796	0.830	0.911	0.923	0.850	0.874	0.937	0.943
ResNet50d	WSIs	0.893	0.895	0.812	0.837	0.909	0.926	0.857	0.869	0.932	0.937
ResNet101d	WSIs	0.893	0.897	0.809	0.827	0.910	0.917	0.857	0.865	0.941	0.947
ResNeXt101	WSIs	0.896	0.905	0.787	0.825	0.941	0.947	0.857	0.873	0.942	0.948
VGG11_bn	WSIs	0.888	0.900	0.786	0.800	0.920	0.943	0.848	0.863	0.933	0.941
VGG19_bn	WSIs	0.903	0.910	0.811	0.821	0.934	0.945	0.868	0.878	0.935	0.938
BEITV2	WSIs	0.892	0.906	0.792	0.803	0.925	0.954	0.853	0.872	0.928	0.952
CoAtNet	WSIs	0.891	0.907	0.779	0.808	0.935	0.950	0.849	0.873	0.934	0.948
CrossViT	WSIs	0.891	0.896	0.793	0.808	0.920	0.930	0.852	0.861	0.951	0.955
DEiT	WSIs	0.886	0.895	0.805	0.811	0.897	0.917	0.849	0.859	0.940	0.947
GCViT	WSIs	0.912	0.918	0.822	0.830	0.948	0.966	0.880	0.888	0.945	0.955
MobileViTV2	WSIs	0.901	0.903	0.805	0.817	0.936	0.948	0.865	0.867	0.947	0.964
SwinV2	WSIs	0.903	0.910	0.797	0.813	0.950	0.965	0.866	0.877	0.933	0.946
ViT	WSIs	0.904	0.902	0.826	0.840	0.951	0.957	0.884	0.895	0.953	0.962

TABLE 3. Results of various deep learning models with imbalanced dataset

	Data	Accuracy		Recall		Precision		F1-score		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
ConvNeXtV2	WSIs	0.935	0.946	0.929	0.940	0.888	0.923	0.908	0.920	0.984	0.989
DenseNet201	WSIs	0.919	0.926	0.901	0.911	0.868	0.888	0.884	0.892	0.969	0.978
EfficientNetV2_b0	WSIs	0.902	0.917	0.894	0.909	0.834	0.865	0.863	0.882	0.967	0.974
EfficientNetV2_b3	WSIs	0.913	0.921	0.900	0.914	0.855	0.871	0.877	0.888	0.972	0.979
MobileNetV3	WSIs	0.914	0.928	0.893	0.919	0.861	0.877	0.877	0.898	0.971	0.976
ResNet50d	WSIs	0.937	0.949	0.923	0.934	0.896	0.924	0.909	0.926	0.977	0.984
ResNet101d	WSIs	0.911	0.921	0.927	0.945	0.834	0.862	0.878	0.892	0.965	0.972
ResNeXt101	WSIs	0.935	0.945	0.909	0.924	0.902	0.929	0.905	0.918	0.977	0.982
VGG11_bn	WSIs	0.922	0.949	0.916	0.936	0.865	0.916	0.890	0.926	0.977	0.986
VGG19_bn	WSIs	0.925	0.939	0.912	0.926	0.876	0.906	0.894	0.912	0.976	0.982
BEITV2	WSIs	0.938	0.950	0.917	0.932	0.904	0.942	0.910	0.927	0.982	0.990
CoAtNet	WSIs	0.939	0.949	0.914	0.923	0.908	0.930	0.911	0.927	0.980	0.986
CrossViT	WSIs	0.929	0.945	0.928	0.939	0.874	0.912	0.900	0.921	0.983	0.989
DEiT	WSIs	0.936	0.944	0.920	0.944	0.897	0.912	0.908	0.920	0.985	0.989
GCViT	WSIs	0.933	0.956	0.928	0.944	0.884	0.929	0.905	0.937	0.983	0.991
MobileViTV2	WSIs	0.914	0.927	0.899	0.909	0.860	0.898	0.878	0.893	0.971	0.977
SwinV2	WSIs	0.942	0.954	0.948	0.954	0.891	0.915	0.919	0.934	0.987	0.993
ViT	WSIs	0.913	0.932	0.893	0.911	0.860	0.897	0.876	0.901	0.973	0.983

TABLE 4. Results of machine learning with balanced dataset (WSIs)

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
Using CNN-based features											
SVM	WSIs	0.914	0.916	0.815	0.816	0.964	0.970	0.883	0.885	0.937	0.940
MLP	WSIs	0.905	0.912	0.807	0.815	0.946	0.969	0.871	0.878	0.935	0.941
LightGBM	WSIs	0.910	0.915	0.822	0.830	0.943	0.954	0.878	0.884	0.924	0.933
Random Forest	WSIs	0.910	0.913	0.810	0.813	0.954	0.960	0.876	0.881	0.945	0.948
Using Transformer-based features											
SVM	WSIs	0.914	0.918	0.821	0.829	0.954	0.965	0.883	0.888	0.940	0.949
MLP	WSIs	0.913	0.926	0.810	0.842	0.964	0.977	0.880	0.900	0.945	0.958
LightGBM	WSIs	0.918	0.922	0.819	0.826	0.968	0.974	0.887	0.894	0.955	0.960
Random Forest	WSIs	0.916	0.921	0.811	0.822	0.971	0.978	0.884	0.892	0.937	0.943

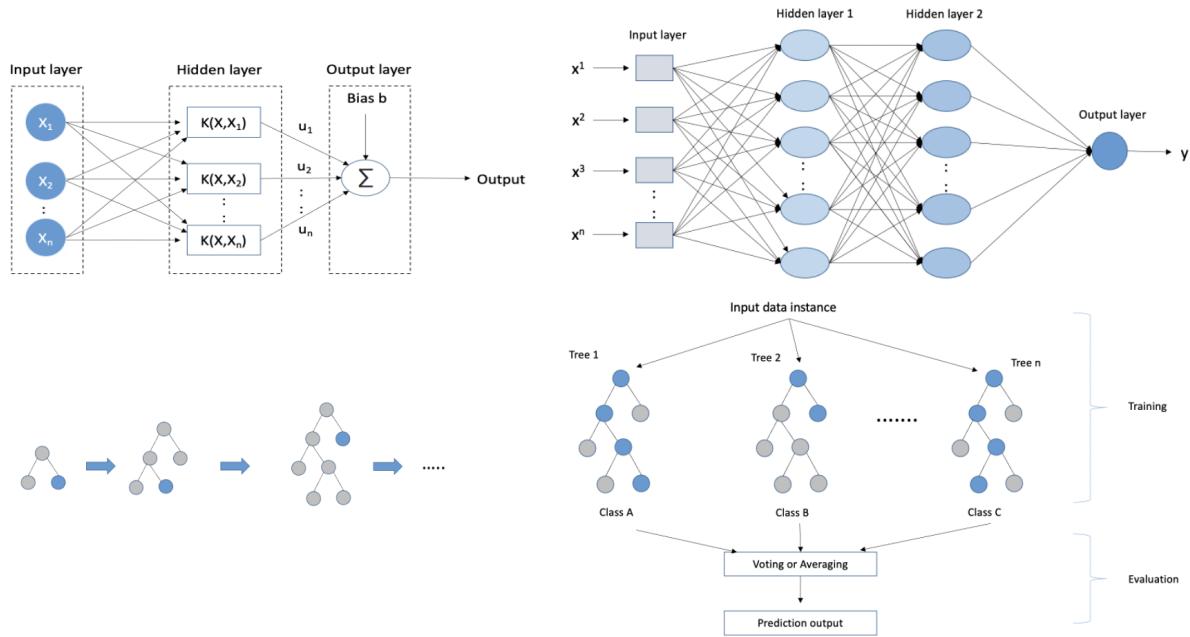


FIGURE 6. The ML classifiers used on ensembled image-based deep learning features in our EFAI framework, namely the SVM, MLP, LightGBM, Random Forest (RF).

TABLE 5. Results of machine learning with imbalanced dataset (WSIs)

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
Using CNN-based features											
SVM	WSIs	0.944	0.955	0.920	0.939	0.916	0.930	0.918	0.935	0.984	0.990
MLP	WSIs	0.946	0.962	0.930	0.948	0.913	0.941	0.921	0.944	0.985	0.993
LightGBM	WSIs	0.950	0.958	0.927	0.932	0.926	0.950	0.927	0.937	0.984	0.985
Random Forest	WSIs	0.949	0.953	0.901	0.905	0.948	0.960	0.924	0.929	0.979	0.982
Using Transformer-based features											
SVM	WSIs	0.950	0.953	0.920	0.927	0.932	0.941	0.926	0.930	0.990	0.990
MLP	WSIs	0.954	0.965	0.926	0.940	0.938	0.956	0.932	0.948	0.989	0.993
LightGBM	WSIs	0.961	0.971	0.920	0.943	0.964	0.972	0.941	0.956	0.991	0.992
Random Forest	WSIs	0.959	0.964	0.937	0.939	0.943	0.955	0.940	0.947	0.991	0.993

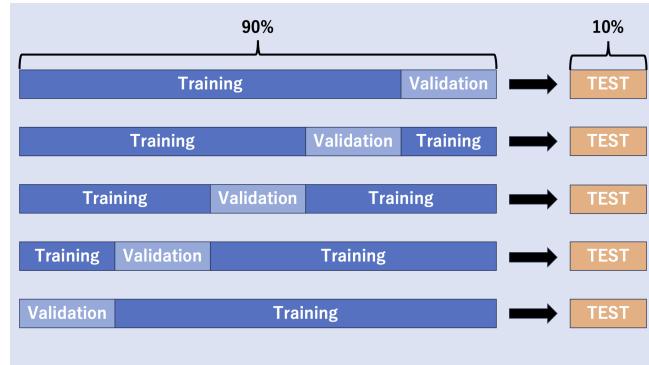


FIGURE 7. Training, validation, and test split used in our experiments.

A. EXPERIMENTS IN DL MODELS (EXPERIMENT 1)

In our study, we initially employed cutting-edge CNN and transformer models to evaluate their performance across the

whole slide image (WSI) database, aiming to establish a performance benchmark. The evaluation was structured into two segments: one involving the balanced dataset and the other the unbalanced dataset, as delineated in Tables 2 and 3 respectively. Upon analysis of the balanced dataset, it was evident that among the CNN architectures, DenseNet201, VGG11_bn, and VGG19_bn demonstrated superior performance. On the transformer side, CoAtNet, GCViT, and SwinV2 provided the best performance. A similar pattern emerged with the unbalanced dataset, where ConvNeXtV2, ResNet50d, and VGG11_bn led the CNN models, and BEiT V2, GCViT, and SwinV2 excelled among the transformer models. Notably, VGG11_bn, GCViT, and SwinV2 consistently showed top performance across both balanced and unbalanced datasets, underscoring their robustness and efficiency. Our evaluation criteria primarily focused on accuracy, alongside recall, precision, F1-score, and AUC scores, to ensure a comprehensive assessment of model capabilities. The

insights garnered from these results have been instrumental in pinpointing the most promising models for further exploration in subsequent experimental phases. This selective approach enables us to concentrate our efforts on the most effective model environments, potentially leveraging techniques such as model ensembling to achieve even greater accuracy. In what follows, we utilized the top 3 models selected from CNNs and Transformers (see Table 1) in both balanced and imbalanced cases.

B. ASSESSING THE ENSEMBLE METHOD (EXPERIMENT 2)

Following the identification of top-performing models, our research progressed to the next experimental phase, focusing on feature extraction from the classification layers of both CNNs and transformers. The aim was to amalgamate these features into a comprehensive feature set for application with various machine learning (ML) classifiers, including Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Light Gradient Boosting Model (LightGBM), and Random Forest. This stage culminated in benchmarking the outcomes to ascertain the most effective pipeline. Feature extraction was performed on an individual patch basis, yielding separate feature sets from each of the three chosen CNNs and transformers. These were then combined to produce an aggregated feature set for each patch, resulting in two consolidated feature sets - one from the CNNs and another from the transformers. These combined feature sets were then subjected to the aforementioned ML classifiers. Analysis of the ensemble results, as detailed in Tables 4 and 5 for both balanced and unbalanced datasets respectively, revealed noteworthy insights. A comparison of these results with those in Tables 2 and 4 highlighted **an enhancement in the average AUC values for both CNN and transformer models, demonstrating the efficacy of the ensemble approach**. Specifically, in the context of a balanced dataset, it was observed that merely excelling in the identification of positive cases (GBM, in our context) was not adequate for achieving an elevated AUC score. This finding underscores the capability of our ensemble strategy to more effectively differentiate between GBM and LGG. When examining the imbalanced dataset through the lens of Tables 3 and 5, a modest improvement in AUC values was recorded, attributed largely to the enhanced classification accuracy of LGGs. This improvement is particularly significant given the dataset's predisposition towards a higher incidence of negatives (LGG) over positives (GBM). In such an imbalanced scenario, achieving a high AUC value alone is not sufficient; accurate classification of GBM becomes crucial. Moreover, the ensemble method was found to enhance the average F1-score for both CNN and transformer models across the board. This improvement is indicative of the ensemble method's capacity to ensure precise classification, even in scenarios where GBM instances are in the minority. Thus, **our findings advocate for the ensemble approach as a potent strategy for navigating the complexities associated with balanced and unbalanced datasets alike**.

C. FUSION OF FEATURES EXTRACTED FROM WSI AND CLINICAL DATASET (EXPERIMENT 3)

Building on the ensembling, we extended our approach by integrating the extracted features from our leading CNN and transformer models with relevant clinical data from patients. This step aimed to evaluate the impact of this integration on performance metrics. We applied this enriched dataset to various ML classifiers, including SVM, MLP, LightGBM, and Random Forest, and subsequently benchmarked the results to identify the most effective pipeline. We utilized hyperparameter optimization of the ML classifiers to obtain the best accuracies overall (see Appendix). The outcomes of this fused dataset application are presented in Tables 6 and 7, corresponding to balanced and unbalanced datasets, respectively. Upon analysis, it was observed that for **both datasets, the transformers outperformed the CNN models in overall performance**. Specifically, **within the CNN category** for the balanced dataset, the **MLP classifier** demonstrated superior efficacy, achieving an impressive 92.7% average accuracy, 83.2% average recall, 97.9% average precision, 90% average F1-score, and 92.7% average AUC score. Meanwhile, among the **transformers**, the **LightGBM classifier** showed the **best result** with an outstanding 93.3% average accuracy, 83.5% average recall, 99.4% average precision, 90.7% average F1-score, and 96% average AUC scores. The scenario of the **unbalanced dataset**, **MLP** once again emerged as **top classifier based on the CNN models**, boasting a 96.4% average accuracy, 92.7% average recall, 96.8% average precision, 94.7% average F1-score, and 99.1% average AUC score. For **transformers**, the **Random Forest classifier** displayed exceptional performance, with 96.3% average accuracy, 94.1% average recall, 95.2% average precision, 94.6% average F1-score, and 99.3% average AUC score. This **analysis reveals that fusing the feature sets from top-performing models with patient clinical data leads to significant improvements in performance metrics**, with transformers generally surpassing **CNNs in effectiveness**. Furthermore, the **results underscore the potential of machine learning classifiers**, particularly **MLP for CNNs and LightGBM and Random Forest for transformers**, in harnessing these combined datasets to **achieve optimal diagnostic accuracy**. Tables 8 through 11 consolidate our experimental outcomes, demonstrating the progression from individual deep learning (DL) model classifications to enhanced results with machine learning (ML) classifiers, and culminating in the augmented performance following the fusion of DL-extracted features with clinical data. Specifically, Tables 8 and 9 focus on CNN models applied to balanced and unbalanced datasets, respectively, whereas Tables 10 and 11 explore transformer models under the same dataset conditions. This comprehensive array of results allows for a detailed comparison of methodologies, showcasing the incremental improvements from standalone DL model performance, through the application of ML classifiers, to the final enhanced outcomes achieved by integrating clinical data into our analysis.

TABLE 6. Results of machine learning with balanced dataset (WSIs+Clinical)

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
Using CNN-based features											
SVM	WSIs	0.927	0.930	0.828	0.831	0.984	0.993	0.899	0.904	0.922	0.925
	Clinical										
MLP	WSIs	0.927	0.932	0.832	0.845	0.979	0.983	0.900	0.908	0.927	0.944
	Clinical										
LightGBM	WSIs	0.920	0.923	0.820	0.825	0.976	0.982	0.891	0.894	0.923	0.939
	Clinical										
Random Forest	WSIs	0.914	0.919	0.817	0.822	0.961	0.968	0.883	0.889	0.947	0.952
	Clinical										
Using Transformer-based features											
SVM	WSIs	0.928	0.936	0.823	0.842	0.993	0.998	0.900	0.912	0.932	0.938
	Clinical										
MLP	WSIs	0.931	0.934	0.830	0.840	0.992	0.995	0.904	0.909	0.944	0.947
	Clinical										
LightGBM	WSIs	0.933	0.936	0.835	0.842	0.994	0.996	0.907	0.913	0.960	0.967
	Clinical										
Random Forest	WSIs	0.918	0.922	0.818	0.823	0.971	0.977	0.888	0.893	0.952	0.956
	Clinical										

TABLE 7. Results of machine learning with imbalanced dataset (WSIs+Clinical)

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
Using CNN-based features											
SVM	WSIs	0.960	0.965	0.931	0.939	0.951	0.959	0.941	0.949	0.992	0.994
	Clinical										
MLP	WSIs	0.964	0.975	0.927	0.943	0.968	0.986	0.947	0.962	0.991	0.997
	Clinical										
LightGBM	WSIs	0.948	0.957	0.904	0.911	0.943	0.962	0.923	0.936	0.975	0.980
	Clinical										
Random Forest	WSIs	0.958	0.962	0.919	0.923	0.958	0.968	0.938	0.943	0.988	0.989
	Clinical										
Using Transformer-based features											
SVM	WSIs	0.953	0.961	0.941	0.948	0.923	0.944	0.932	0.943	0.989	0.991
	Clinical										
MLP	WSIs	0.958	0.978	0.944	0.964	0.937	0.977	0.940	0.968	0.991	0.995
	Clinical										
LightGBM	WSIs	0.963	0.970	0.921	0.943	0.970	0.987	0.945	0.955	0.991	0.993
	Clinical										
Random Forest	WSIs	0.963	0.969	0.941	0.944	0.952	0.964	0.946	0.954	0.993	0.994
	Clinical										

TABLE 8. Single deep learning model vs Ensemble of top 3 models with WSIs vs Ensemble of top 3 models with WSIs+Clinical : Using CNN-based models and balanced dataset

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
Single deep learning model vs Ensemble of top 3 models with WSIs											
DenseNet201	WSIs	0.894	0.912	0.803	0.789	0.933	0.949	0.855	0.881	0.898	0.921
VGG19_bn	WSIs	0.903	0.910	0.811	0.821	0.934	0.945	0.868	0.878	0.933	0.941
VGG11_bn	WSIs	0.888	0.900	0.786	0.800	0.920	0.943	0.848	0.863	0.935	0.938
SVM	WSIs	0.914	0.916	0.815	0.816	0.964	0.970	0.883	0.885	0.937	0.940
MLP	WSIs	0.905	0.912	0.807	0.815	0.946	0.969	0.871	0.878	0.935	0.941
LightGBM	WSIs	0.910	0.915	0.822	0.830	0.943	0.954	0.878	0.884	0.924	0.933
Random Forest	WSIs	0.910	0.913	0.810	0.813	0.954	0.960	0.876	0.881	0.945	0.948
Ensemble of top 3 models with WSIs+Clinical : Using CNN-based models											
SVM	WSIs	0.927	0.930	0.828	0.831	0.984	0.993	0.899	0.904	0.922	0.925
MLP	WSIs	0.927	0.932	0.832	0.845	0.979	0.983	0.900	0.908	0.927	0.944
LightGBM	WSIs	0.920	0.923	0.820	0.825	0.976	0.982	0.891	0.894	0.923	0.939
Random Forest	WSIs	0.914	0.919	0.817	0.822	0.961	0.968	0.883	0.889	0.947	0.952

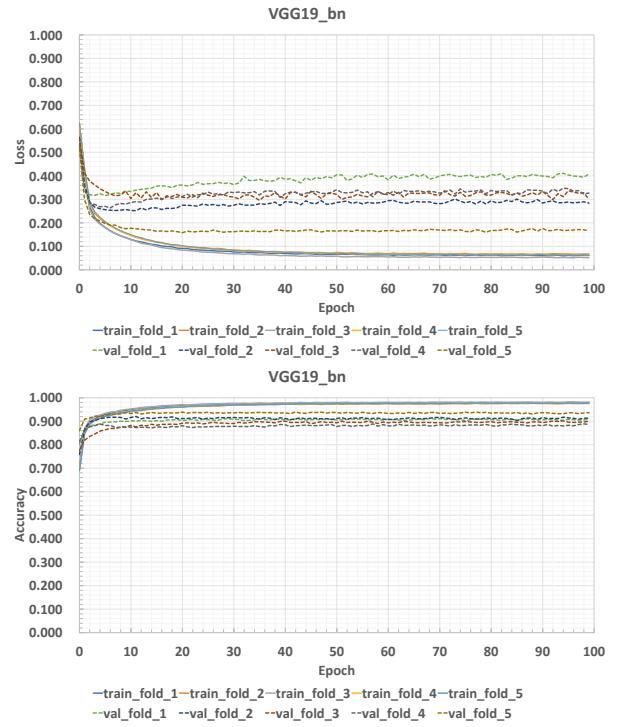
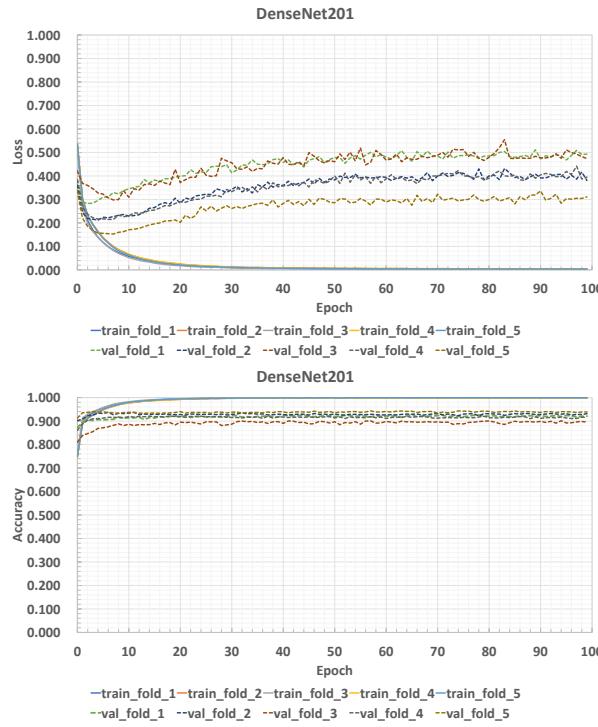


FIGURE 8. DenseNet201 loss/accuracy curve with balanced dataset

FIGURE 9. VGG19_bn loss/accuracy curve with balanced dataset

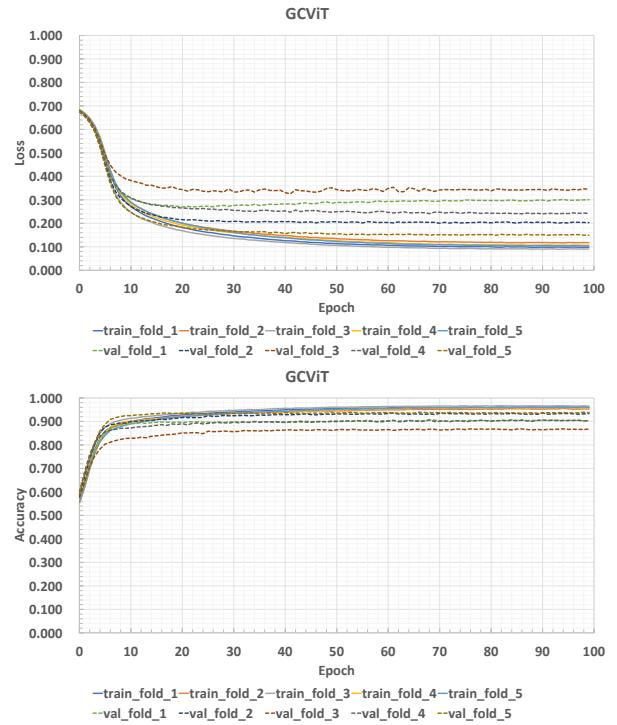
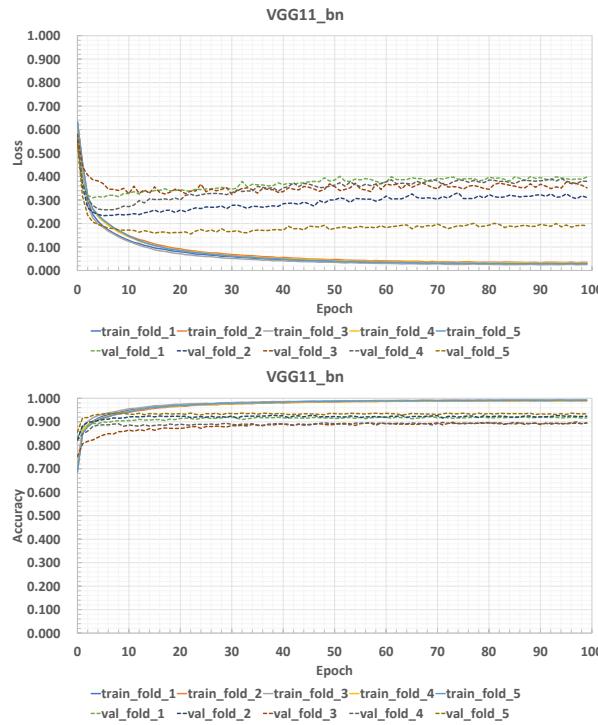


FIGURE 10. VGG11_bn loss/accuracy curve with balanced dataset

FIGURE 11. GCViT loss/accuracy curve with balanced dataset

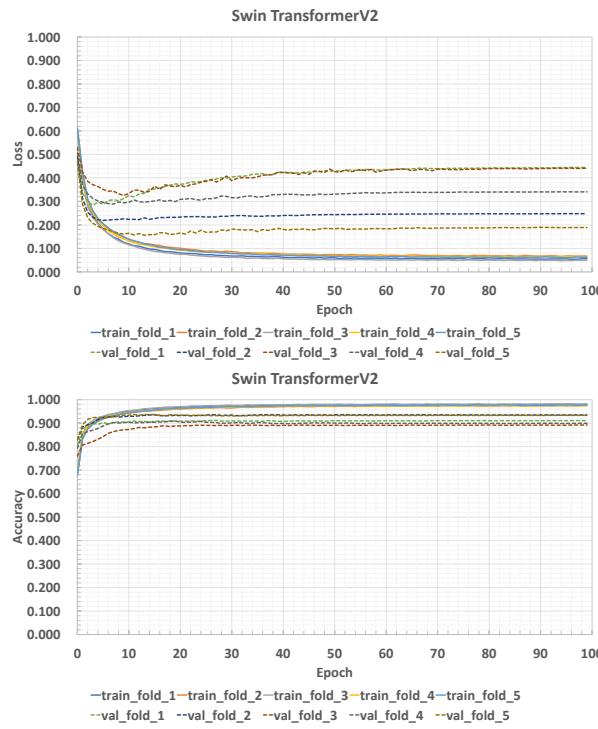


FIGURE 12. Swin TransformerV2 loss/accuracy curve with balanced dataset

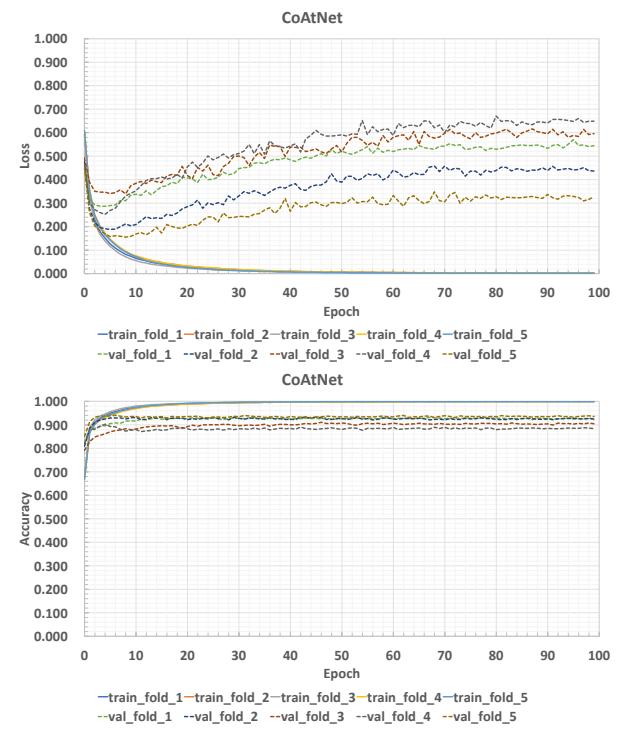


FIGURE 13. CoAtNet loss/accuracy curve with balanced dataset

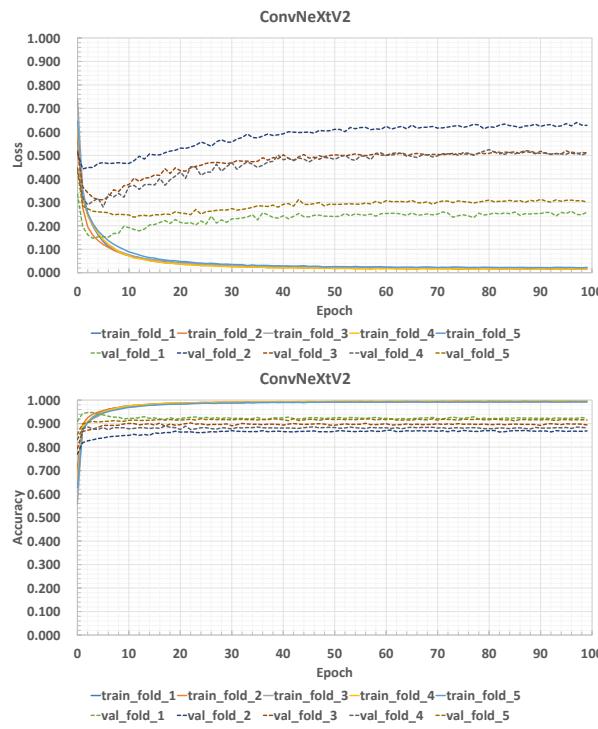


FIGURE 14. ConvNeXtV2 loss/accuracy curve with imbalanced dataset

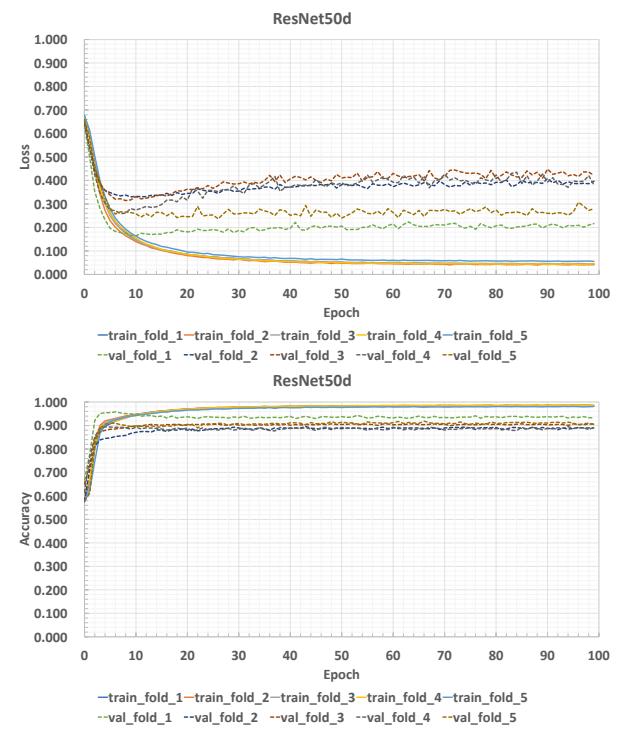


FIGURE 15. ResNet50d loss/accuracy curve with imbalanced dataset

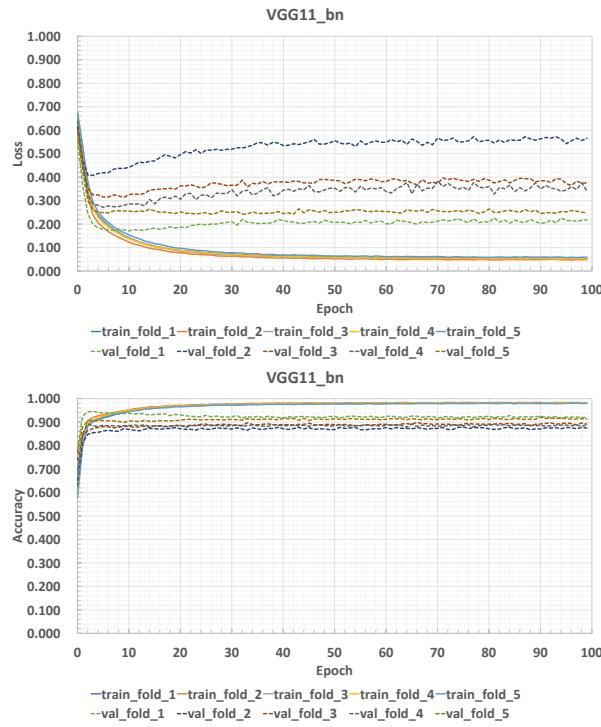


FIGURE 16. VGG11_bn loss/accuracy curve with imbalanced dataset

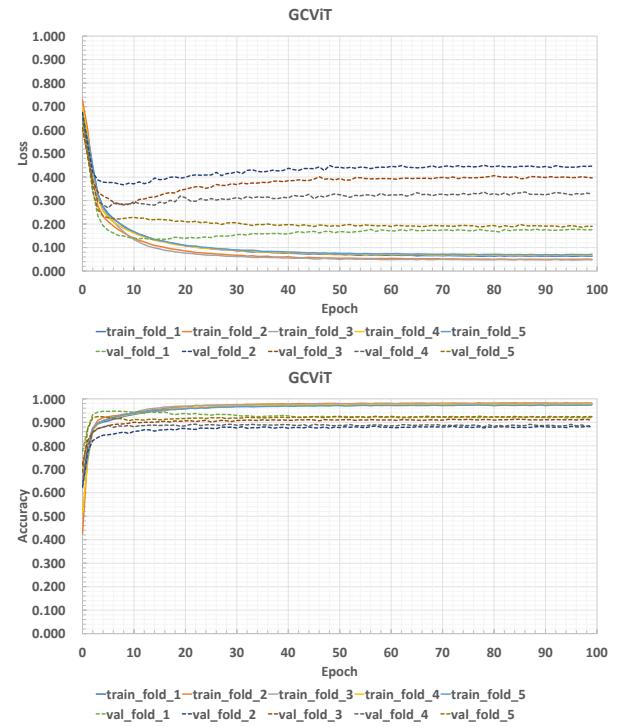


FIGURE 17. GCViT loss/accuracy curve with imbalanced dataset

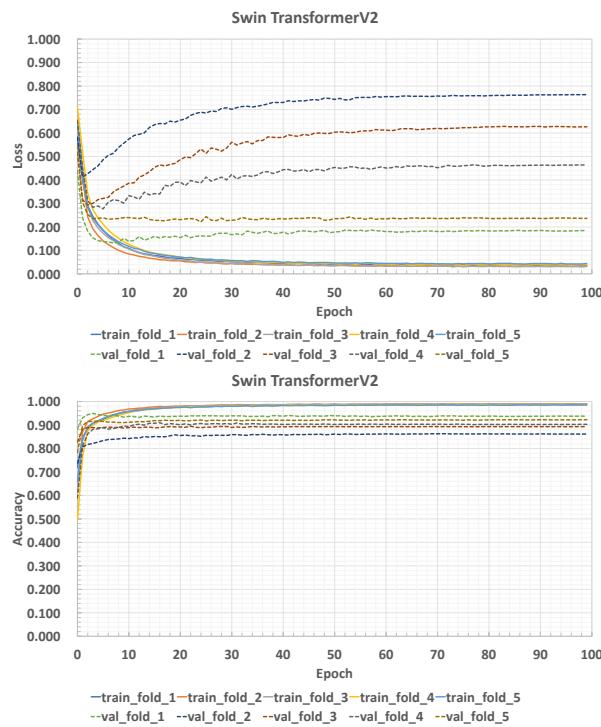


FIGURE 18. Swin TransformerV2 loss/accuracy curve with imbalanced dataset

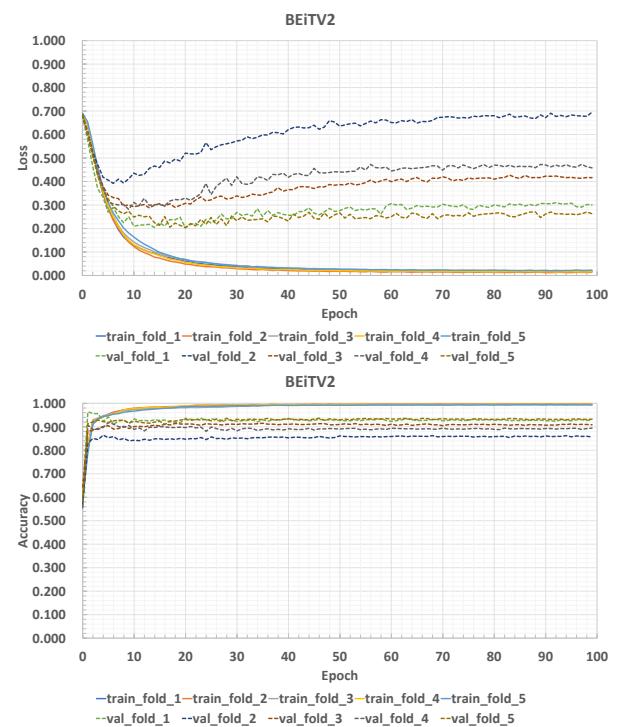


FIGURE 19. BEiT2 loss/accuracy curve with imbalanced dataset

TABLE 9. Single deep learning model vs Ensemble of top 3 models with WSIs vs Ensemble of top 3 models with WSIs+Clinical : Using Transformer-based models and balanced dataset

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
GCViT	WSIs	0.912	0.918	0.822	0.830	0.948	0.966	0.880	0.888	0.945	0.955
SwinV2	WSIs	0.903	0.910	0.797	0.813	0.950	0.965	0.866	0.877	0.933	0.946
CoAtNet	WSIs	0.891	0.907	0.779	0.808	0.935	0.950	0.849	0.873	0.934	0.948
SVM	WSIs	0.914	0.918	0.821	0.829	0.954	0.965	0.883	0.888	0.940	0.949
MLP	WSIs	0.913	0.926	0.810	0.842	0.964	0.977	0.880	0.900	0.945	0.958
LightGBM	WSIs	0.918	0.922	0.819	0.826	0.968	0.974	0.887	0.894	0.955	0.960
Random Forest	WSIs	0.916	0.921	0.811	0.822	0.971	0.978	0.884	0.892	0.937	0.943
SVM	WSIs Clinical	0.928	0.936	0.823	0.842	0.993	0.998	0.900	0.912	0.932	0.938
MLP	WSIs Clinical	0.931	0.934	0.830	0.840	0.992	0.995	0.904	0.909	0.944	0.947
LightGBM	WSIs Clinical	0.933	0.936	0.835	0.842	0.994	0.996	0.907	0.913	0.960	0.967
Random Forest	WSIs Clinical	0.918	0.922	0.818	0.823	0.971	0.977	0.888	0.893	0.952	0.956

TABLE 10. Single deep learning model vs Ensemble of top 3 models with WSIs vs Ensemble of top 3 models with WSIs+Clinical : Using CNN-based models and imbalanced dataset

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
ResNet50d	WSIs	0.937	0.949	0.923	0.934	0.896	0.924	0.909	0.926	0.977	0.984
VGG11_bn	WSIs	0.922	0.949	0.916	0.936	0.865	0.916	0.890	0.926	0.977	0.986
ConvNeXtV2	WSIs	0.935	0.946	0.929	0.940	0.888	0.923	0.908	0.920	0.984	0.989
SVM	WSIs	0.944	0.955	0.920	0.939	0.916	0.930	0.918	0.935	0.984	0.990
MLP	WSIs	0.946	0.962	0.930	0.948	0.913	0.941	0.921	0.944	0.985	0.993
LightGBM	WSIs	0.950	0.958	0.927	0.932	0.926	0.950	0.927	0.937	0.984	0.985
Random Forest	WSIs	0.949	0.953	0.901	0.905	0.948	0.960	0.924	0.929	0.979	0.982
SVM	WSIs Clinical	0.960	0.965	0.931	0.939	0.951	0.959	0.941	0.949	0.992	0.994
MLP	WSIs Clinical	0.964	0.975	0.927	0.943	0.968	0.986	0.947	0.962	0.991	0.997
LightGBM	WSIs Clinical	0.948	0.957	0.904	0.911	0.943	0.962	0.923	0.936	0.975	0.980
Random Forest	WSIs Clinical	0.958	0.962	0.919	0.923	0.958	0.968	0.938	0.943	0.988	0.989

TABLE 11. Single deep learning model vs Ensemble of top 3 models with WSIs vs Ensemble of top 3 models with WSIs+Clinical : Using Transformer-based models and imbalanced dataset

	Data	Accuracy		Recall		Precision		F1		AUC	
		Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
GCViT	WSIs	0.933	0.956	0.928	0.944	0.884	0.929	0.905	0.937	0.983	0.991
SwinV2	WSIs	0.942	0.954	0.948	0.954	0.891	0.915	0.919	0.934	0.987	0.993
BEiTV2	WSIs	0.938	0.950	0.917	0.932	0.904	0.942	0.910	0.927	0.982	0.990
SVM	WSIs	0.950	0.953	0.920	0.927	0.932	0.941	0.926	0.930	0.990	0.990
MLP	WSIs	0.954	0.965	0.926	0.940	0.938	0.956	0.932	0.948	0.989	0.993
LightGBM	WSIs	0.961	0.971	0.920	0.943	0.964	0.972	0.941	0.956	0.991	0.992
Random Forest	WSIs	0.959	0.964	0.937	0.939	0.943	0.955	0.940	0.947	0.991	0.993
SVM	WSIs Clinical	0.953	0.961	0.941	0.948	0.923	0.944	0.932	0.943	0.989	0.991
MLP	WSIs Clinical	0.958	0.978	0.944	0.964	0.937	0.977	0.940	0.968	0.991	0.995
LightGBM	WSIs Clinical	0.963	0.970	0.921	0.943	0.970	0.987	0.945	0.955	0.991	0.993
Random Forest	WSIs Clinical	0.963	0.969	0.941	0.944	0.952	0.964	0.946	0.954	0.993	0.994

D. ASSESSING THE PROPOSED MULTIMODAL ANALYSIS

The experimental results on whole slide images (WSIs) alone and in combination with clinical data are presented in Tables 8 and 9, which show the outcomes for the top three CNNs and transformers, respectively, on balanced datasets. These tables illustrate that combining WSIs with clinical data (**WSI + Clinical**) consistently yields higher F1-scores and accuracy compared to using WSIs alone, for both CNNs and transformers. Tables 10 and 11 extend these comparisons to imbalanced datasets for the top three CNNs and transformers, respectively. The fused datasets demonstrate notably better F1-scores and accuracy for CNNs and show marginal improvements for transformers. Across both balanced and imbalanced datasets, the performance metrics are enhanced when WSIs are combined with clinical data. Classification results for the top three ensembled models using only WSIs are detailed in Tables 4 and 5 for balanced and imbalanced datasets, respectively. These results indicate that transformers generally outperform CNNs. However, Tables 6 and 7, which present classification outcomes for both CNNs and transformers using combined WSIs and clinical data, show no significant performance differential between the two model types.

The leading CNN models from our experiments using only WSIs include DenseNet201, VGG11_bn, VGG19_bn, ResNet50d, and ConvNeXtV2. The top-performing transformer models are GCViT, Swin TransformerV2, CoAtNet, and BEiT2. Notably, VGG11_bn is identified as the most efficient variant among the VGG models. Similarly, GCViT offers a more computationally efficient alternative to the traditional Vision Transformer (ViT). Ensembling these top models has shown to enhance classification performance, validating the efficacy of our proposed ensemble method, especially when incorporating lightweight models. VGG11_bn, GCViT, and Swin TransformerV2 stood out among the top models for both balanced and imbalanced datasets. While GCViT and ConvNeXtV2 are still seldom used in pathology-based models, this research underscores their effectiveness in leveraging global context self-attention and standard local self-attention, which are key features of GCViT.

In Figure 8 we showed the loss and accuracy curves for DenseNet201. The loss curves show that the model is learning with training loss converging steadily. Validation loss appears relatively stable yet fluctuates, hinting at possible improvements through hyperparameter tuning. The consistency across all folds indicates the model's reliable performance on the dataset, and an early stopping strategy around 40-60 epochs could be beneficial to prevent overfitting. The accuracy curves show a rapid learning rate initially during training and a flattening curve after 20 - 30 epochs. Validation accuracy shows less improvement and stabilizes at a lower level than the training accuracy. Overall uniformity across the five folds reflects consistent model behavior, reinforcing the robustness of the model across different data subsets. Figure 9 features the VGG19_bn model's performance, with loss curves converging smoothly around 0.2 after about 30 epochs,

indicative of the model's stable learning from the data. Accuracy curves plateau early near the 0.9 mark, suggesting that most learning occurs in the initial epochs. Across all five folds, the consistency in both loss and accuracy metrics demonstrates the model's robustness and reliability on the balanced dataset. Figure 10 presents the loss and accuracy graphs for the VGG11_bn model trained on a balanced dataset over 100 epochs. The loss curves demonstrate convergence below 0.3, signifying effective learning, while the accuracy curves stabilize close to 0.9, reflecting high predictive performance. The parallelism of the curves across the five validation folds indicates consistent model behavior and generalization capabilities. Figure 11 shows the loss and accuracy performance of the GCViT model, where the loss curves flatten out just above 0.1, indicating an efficient learning process. The accuracy curves ascend swiftly and plateau near the 0.9 threshold, suggesting early achievement of high predictive accuracy. Uniformity across the five folds reflects the model's robust generalization across different subsets of the data. Figure 12 depicts the training journey of the Swin Transformer V2 model, with the loss curves leveling out around 0.2, pointing to a stable learning trajectory. Accuracy for both training and validation demonstrates a plateau approximately at 0.9, hinting at the model's quick adaptation and subsequent consistent performance. The close tracking of training and validation lines across five folds suggests a high level of model reliability and transferability across the balanced dataset. Figure 13 presents the CoAtNet model's loss and accuracy curves over 100 epochs. The loss curves indicate initial rapid learning, stabilizing around 0.3, with the validation loss exhibiting some oscillations, suggesting potential areas for refinement. The accuracy curves show the model achieving high performance levels, with both training and validation accuracies nearing the 0.9 mark and displaying uniformity across different data folds, affirming the model's effectiveness and consistency. Figure 14 showcases the ConvNeXtV2 model's loss and accuracy curves on an imbalanced dataset over 100 epochs. The loss curves reveal a gradual decrease, plateauing near 0.4 for validation, which might indicate the need for model tuning to better address the dataset imbalance. Accuracy graphs suggest the model reaches near-peak performance quickly, maintaining above 0.9 throughout the training, which is consistently mirrored in the validation accuracy across all folds. Figure 15 features the ResNet50d model's loss and accuracy curves on an imbalanced dataset across 100 epochs. The loss curves indicate a consistent decline, with training loss dipping just below 0.1 and validation loss plateauing around 0.3, hinting at the model's ability to learn despite data imbalance. The accuracy plots climb quickly to maintain a level close to 0.9, showcasing the model's strong predictive capabilities, which remain steady across all data folds. Figure 16 illustrates the VGG11_bn model's loss and accuracy curves for an imbalanced dataset over 100 epochs. The loss curves show training loss declining to around 0.2, with validation loss exhibiting some variance, settling near 0.35, suggesting room for improved model calibration. The

accuracy curves for both training and validation rise swiftly and plateau above 0.85, reflecting the model's adeptness at handling imbalanced data, evidenced by the consistent pattern across all folds. Figure 17 depicts the performance curves for the GCViT model on an imbalanced dataset over the course of 100 epochs. The loss graph indicates that the training loss decreases effectively, reaching a level just below 0.2, while the validation loss exhibits some minor fluctuations, stabilizing close to 0.2. The accuracy curves ascend quickly and level off, maintaining around 0.95, suggesting the model's high efficiency in classification tasks despite the imbalanced nature of the dataset. The consistency in performance across all five folds suggests a reliable and stable model. Figure 18 shows the learning dynamics of the Swin Transformer V2 model trained on an imbalanced dataset. The loss curves descend significantly in early epochs, with training loss reaching around 0.2 and validation loss showing a slight oscillation but generally remaining under 0.25. The accuracy curves demonstrate a rapid increase, plateauing at approximately 0.95, which indicates the model's robust capability to handle the imbalanced data across the varying folds. The graphs reflect a high level of consistency, reinforcing the model's stability and generalizability.

Figure 19 depicts the training process of the BEiT V2 model on an imbalanced dataset, as demonstrated over 100 epochs. The loss curves reflect a rapid decrease initially, with training loss stabilizing just below 0.2, while validation loss levels off slightly higher, suggesting the model is learning effectively but may benefit from further tuning to address validation set challenges. Accuracy plots exhibit a swift ascent, stabilizing at high levels close to 0.95, revealing the model's proficient learning from the imbalanced dataset across all data folds. The consistent trends across the folds suggest the model's robust generalization.

Comparing the classification results of the proposed method using only WSIs, all metrics are higher in the case of Transformer-based features than in the case of CNN-based features. The Transformer-based models (GCViT, Swin TransformerV2, CoAtNet, and BEiT V2) used in this research for feature extraction of histopathological images contain an internal attention mechanism. Therefore, the feature vectors obtained from these models consider the importance of the attention weight, and it is considered that the Transformer-based features performed better than the CNN-based features. The results of this research prove the usefulness of using Transformer-based models as a backbone for pathology-related tasks. Next, the classification performance of machine learning models that classify using only histopathological images or histopathological images and clinical data features is compared. This comparison focuses on experiments using transformer-based features. First, in the case of the balanced dataset, the LightGBM outperforms the other models in the average Accuracy, F1-score, and AUC in the experiment using only histopathological image features. Furthermore, in the experiment using features from histopathological images and clinical data, the LightGBM has higher average

Accuracy, F1-score, and AUC than the other compared models. Hence, overall LightGBM showed the best classification performance when using the balanced dataset. On the other hand, in the case of the imbalanced dataset, LightGBM and RandomForest showed higher average Accuracy, F1-score, and AUC than the other models in the experiment using only histopathological image features. Furthermore, LightGBM and RandomForest outperform the other models in the average Accuracy, F1-score, and AUC in the experiment using features from histopathological images and clinical data. Therefore, in the case of the imbalanced dataset, LightGBM and RandomForest showed the best classification performance. The results of this extensive comparisons proves the versatility of the decision tree-based classifiers [45], as they performed better than the other models in each of the two dataset experiments. In the data input to the machine learning model in this research, the features of the histopathological images are complex because they are extracted from multiple deep learning models. The complexity is further increased by the addition of features from clinical data. We consider that LightGBM and RandomForest are machine learning models based on ensemble learning, which combines multiple decision trees, and thus can be applied to complex input features.

Given the large-scale nature of whole slide images (WSIs) and multimodal data, computational efficiency is a key factor in the implementation of our proposed approach. To mitigate computational overhead, we employ pre-trained CNN and Transformer models for feature extraction, eliminating the need for full-scale deep learning model training. Additionally, Principal Component Analysis (PCA) is applied to reduce the dimensionality of extracted features, ensuring that only the most relevant information is retained while significantly reducing processing time. For classification, we utilize efficient machine learning models such as LightGBM and Random Forest, which have lower computational complexity compared to deep neural networks. These optimizations ensure that the proposed ensemble fusion AI (EFAI) framework remains computationally feasible and scalable for real-world clinical applications.

While our approach achieves high classification accuracy, its real-world adoption requires better model transparency for clinicians. The use of machine learning classifiers such as LightGBM and Random Forest in the final prediction stage helps interpretability by providing feature importance scores, making the decision-making process more understandable. Additionally, Transformer-based models inherently use attention mechanisms, which indicate the most relevant image regions contributing to predictions. Future work will focus on further improving interpretability by developing clinician-friendly visualization techniques and assessing the model's applicability in real-world clinical settings.

IV. CONCLUSION

Brain glioma subtype classification from histopathological whole slide images (WSIs) is a challenging problem. In this work, we considered classification of glioma subtypes

TABLE 12. List of optimal hyperparameters for machine learning classifiers for Balanced dataset.

Feature Type	Classifier's Name	Hyperparameters
Ensembled CNN-based features	SVM	kernel: linear, C: 1e-3
	MLP	hidden layer sizes: 4, activation: ReLU, solver: Adam, learning rate: 1e-3
	LightGBM	boosting type: gdbt, n_estimator: 200, learning rate: 1e-2
	Random Forest	n_estimator: 200
Fused CNN-based features and clinical data	SVM	kernel: rbf, gamma: scale, C: 10
	MLP	hidden layer sizes: 3, activation: ReLU, solver: Adam, learning rate: 1e-2
	LightGBM	boosting type: gdgb, n_estimator: 1000, learning rate: 1e-3
	Random Forest	n_estimator: 200
Ensembled Transformer-based features	SVM	kernel: rbf, gamma: auto, C: 1e-1
	MLP	hidden layer sizes: 3, activation: ReLU, solver: Adam, learning rate: 1e-2
	LightGBM	boosting type: gdgb, n_estimator: 1000, learning rate: 1e-2
	Random Forest	n_estimator: 20
Transformer-based features and clinical data	SVM	kernel: rbf, gamma: auto, C: 10
	MLP	hidden layer sizes: 5, activation: ReLU, solver: Adam, learning rate: 1e-4
	LightGBM	boosting type: gdgb, n_estimator: 400, learning rate: 1e-2
	Random Forest	n_estimator: 800

TABLE 13. List of alternative hyperparameters for machine learning classifiers for Imbalanced dataset.

Feature Type	Classifier's Name	Hyperparameters
Ensembled CNN-based features	SVM	kernel: rbf, gamma: scale, C: 10
	MLP	hidden layer sizes: 3, activation: ReLU, solver: Adam, learning rate: 1e-4
	LightGBM	boosting type: gdgb, n_estimator: 80, learning rate: 1e-1
	Random Forest	n_estimator: 10
Fused CNN-based features and clinical data	SVM	kernel: rbf, gamma: scale, C: 1
	MLP	hidden layer sizes: 5, activation: ReLU, solver: Adam, learning rate: 1e-3
	LightGBM	boosting type: gdgb, n_estimator: 10, learning rate: 1e-1
	Random Forest	n_estimator: 100
Ensembled Transformer-based features	SVM	kernel: rbf, gamma: auto, C: 1e-2
	MLP	hidden layer sizes: 2, activation: ReLU, solver: Adam, learning rate: 1e-4
	LightGBM	boosting type: gdgb, n_estimator: 400, learning rate: 1e-3
	Random Forest	n_estimator: 20
Fused Transformer-based features and clinical data	SVM	kernel: linear, C: 10
	MLP	hidden layer sizes: 3, activation: ReLU, solver: Adam, learning rate: 1e-2
	LightGBM	boosting type: gdgb, n_estimator: 400, learning rate: 1e-3
	Random Forest	n_estimator: 80

based on histopathological images and clinical data using multiple state-of-the-art artificial intelligence (AI) models. Our proposed ensemble fusion AI (EFAI) methodology uses features of histopathological images extracted by multiple DL models and clinical data to classify glioma subtypes based on machine learning models. We benchmarked various convolutional neural networks (CNNs), and Transformers with both balanced and imbalanced datasets consisting of histopathological images. **Based on the initial experiments of this benchmarking we selected top models in both CNNs and Transformers and combined them into a global average pooling layer fusion wherein salient features from these models are aggregated for obtaining overall better classification accuracy.** In experiments using only histopathological images of glioma patients, our proposed multimodal ensemble fusion AI approach significantly outperformed the classification results in terms of using any single top-performing CNN or Transformers DL models. Furthermore, in experiments using both histopathological images and clinical data of patients with glioma, the proposed ensembling fusion approach significantly outperformed DL models using only histopathological images. Finally, our experimental results indicate that our proposed multimodal ensemble fusion can be useful in using multiple top-performing deep learning models and can be of potential use in classification of glioma subtypes [46], [47]. Future works include the use of other fusion techniques including attention- and multiple instance-based models [48], [49] that can leverage multiple modalities in a complementary way and fusing other DL models.

CONFLICT OF INTEREST

None.

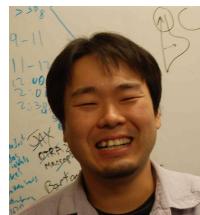
APPENDIX

We experimented with different hyperparameter settings for the machine learning classifiers, adapting them to both balanced and imbalanced datasets, the parameter combinations are given in Tables 12 and 13 respectively. These optimal parameters were tuned to obtain the best average accuracy overall.

REFERENCES

- [1] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "Review the cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemporary Oncology/Współczesna Onkologia*, vol. 2015, no. 1, pp. 68–77, 2015.
- [2] M. Weller, W. Wick, K. Aldape, M. Brada, M. Berger, S. M. Pfister, R. Nishikawa, M. Rosenthal, P. Y. Wen, R. Stupp *et al.*, "Glioma," *Nature Reviews Disease Primers*, vol. 1, no. 1, pp. 1–18, 2015.
- [3] D. N. Louis, A. Perry, G. Reifenberger, A. Von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, Ö. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 world health organization classification of tumors of the central nervous system: a summary," *Acta Neuropathologica*, vol. 131, pp. 803–820, 2016.
- [4] A. Yonekura, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, and H. Takase, "Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network," *Biomedical Engineering Letters*, vol. 8, no. 3, pp. 321–327, 2018.
- [5] J.-P. Redlich, F. Feuerhake, J. Weis, N. S. Schaadt, S. Teuber-Hanselmann, C. Buck, S. Luttmann, A. Eberle, S. Nikolin, A. Appenzeller *et al.*, "Applications of artificial intelligence in the analysis of histopathology images of gliomas: a review," *npj Imaging*, vol. 2, no. 1, p. 16, 2024.
- [6] F. Hanif, M. Muzaffar, K. Perveen, S. M. Malhi, and S. U. Simjee, "Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment," *Asian Pacific Journal of Cancer prevention*, vol. 18, no. 1, p. 3, 2017.
- [7] P. Chang, J. Grinband, B. Weinberg, M. Bardis, M. Khy, G. Cadena, M.-Y. Su, S. Cha, C. Filippi, D. Bota *et al.*, "Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas," *American Journal of Neuroradiology*, vol. 39, no. 7, pp. 1201–1207, 2018.
- [8] Y. Matsui, T. Maruyama, M. Nitta, T. Saito, S. Tsuzuki, M. Tamura, K. Kusuda, Y. Fukuya, H. Asano, T. Kawamata *et al.*, "Prediction of lower-grade glioma molecular subtypes using deep learning," *Journal of neuro-oncology*, vol. 146, pp. 321–327, 2020.
- [9] L. Jin, F. Shi, Q. Chun, H. Chen, Y. Ma, S. Wu, N. F. Hameed, C. Mei, J. Lu, J. Zhang *et al.*, "Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers," *Neuro-oncology*, vol. 23, no. 1, pp. 44–52, 2021.
- [10] R. Nakagaki, S. S. Debsarkar, H. Kawanaka, B. Aronow, and V. B. S. Prasath, "Deep learning-based idh1 gene mutation prediction using histopathological imaging and clinical data," *Computers in Biology and Medicine*, vol. 179, no. 108902, pp. –, 2024.
- [11] M. K. K. Niazi, A. V. Parwani, and M. N. Gurcan, "Digital pathology and artificial intelligence," *The Lancet Oncology*, vol. 20, no. 5, pp. e253–e261, 2019.
- [12] T. Hayakawa, V. B. S. Prasath, H. Kawanaka, B. J. Aronow, and S. Tsuruoka, "Computational nuclei segmentation methods in digital pathology : A survey," *Archives of Computational Methods in Engineering*, vol. 28, no. 1, pp. 1–13, 2021.
- [13] C. Zhou, Y. Jin, Y. Chen, S. Huang, R. Huang, Y. Wang, Y. Zhao, Y. Chen, L. Guo, and J. Liao, "Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101861, 2021.
- [14] S. Banerji and S. Mitra, "Deep learning in histopathology: A review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 1, p. e1439, 2022.
- [15] B. A. Kumar and N. K. Misra, "Colon cancer classification and detection by novel cmnv2 model and methods of deep learning," *Neural Computing and Applications*, pp. 1–17, 2024.
- [16] A. H. Song, G. Jaume, D. F. Williamson, M. Y. Lu, A. Vaidya, T. R. Miller, and F. Mahmood, "Artificial intelligence for digital and computational pathology," *Nature Reviews Bioengineering*, vol. 1, no. 12, pp. 930–949, 2023.
- [17] S. Im, J. Hyeon, E. Rha, J. Lee, H.-J. Choi, Y. Jung, and T.-J. Kim, "Classification of diffuse glioma subtype from clinical-grade pathological images using deep transfer learning," *Sensors*, vol. 21, no. 10, p. 3500, 2021.
- [18] J. Yang, J. Ju, L. Guo, B. Ji, S. Shi, Z. Yang, S. Gao, X. Yuan, G. Tian, Y. Liang *et al.*, "Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 333–342, 2022.
- [19] S. Rathore, T. Niazi, M. A. Iftikhar, and A. Chaddad, "Glioma grading via analysis of digital pathology images using machine learning," *Cancers*, vol. 12, no. 3, p. 578, 2020.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [21] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2021.
- [22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [23] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [24] Z.-H. Zhou, *Machine learning*. Springer nature, 2021.
- [25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [28] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16 133–16 142.
- [29] A. Hatamizadeh, H. Yin, G. Heinrich, J. Kautz, and P. Molchanov, "Global context vision transformers," in *International Conference on Machine Learning*, 2023, pp. 12 633–12 646.
- [30] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong *et al.*, "Swin transformer v2: Scaling up capacity and resolution," in *IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 009–12 019.
- [31] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "Coatnet: Marrying convolution and attention for all data sizes," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3965–3977, 2021.
- [32] Z. Peng, L. Dong, H. Bao, Q. Ye, and F. Wei, "Beit v2: Masked image modeling with vector-quantized visual tokenizers," *arXiv preprint arXiv:2208.06366*, 2022.
- [33] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning*. PMLR, 2021, pp. 10 096–10 106.
- [34] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.
- [35] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492–1500.
- [36] C.-F.R. Chen, Q. Fan, and R. Panda, "Crossvit: Cross-attention multi-scale vision transformer for image classification," in *IEEE/CVF international conference on computer vision*, 2021, pp. 357–366.
- [37] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *International Conference on Machine Learning*, 2021, pp. 10 347–10 357.
- [38] S. Mehta and M. Rastegari, "Separable self-attention for mobile vision transformers," *arXiv preprint arXiv:2206.02680*, 2022.
- [39] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.
- [40] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [41] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [42] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [43] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," *Artificial intelligence*, vol. 137, no. 1-2, pp. 239–263, 2002.
- [44] Y. Yang, H. Lv, and N. Chen, "A survey on ensemble learning under the era of deep learning," *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5545–5589, 2023.

- [45] G. Murtaza, A. W. A. Wahab, G. Raza, and L. Shuib, "A tree-based multiclassification of breast tumor histopathology images through deep learning," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101870, 2021.
- [46] K. Fukuma, H. Kawanaka, S. Prasath, B. J. Aronow, and H. Takase, "Feature extraction and disease stage classification for glioma histopathology images," in *IEEE 17th International Conference on e-Health Networking, Applications and Services (HealthCom)*, Boston, USA, October 2015, pp. 429–430.
- [47] A. Yonekura, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, and H. Takase, "Glioblastoma multiforme tissue histopathology images based disease stage classification with deep CNN," in *IEEE 6th International Conference on Informatics, Electronics & Vision (ICIEV)*, Himeji, Hyogo, Japan, September 2017.
- [48] O. Ester, F. Hörst, C. Seibold, J. Keyl, S. Ting, N. Vasileiadis, J. Schmitz, P. Ivanyi, V. Grünwald, J. H. Bräsen *et al.*, "Valuing vicinity: Memory attention framework for context-based semantic segmentation in histopathology," *Computerized Medical Imaging and Graphics*, vol. 107, p. 102238, 2023.
- [49] M. Gadermayr and M. Tschuchnig, "Multiple instance learning for digital pathology: A review of the state-of-the-art, limitations & future potential," *Computerized Medical Imaging and Graphics*, p. 102337, 2024.



HIROHARU KAWANAKA (Member, IEEE) received a Dr. Eng. degree from Mie University Graduate School of Engineering in 2004. After graduation, he entered Mie University Graduate School of Medicine to study Medical Informatics and Medical/Clinical Information Systems. In 2004, he established a company, "Medical Engineering Institute, Inc." to develop a DWH for clinical use. More than 120 large-scale hospitals use this system to analyze clinical data archived in the EHR for clinical studies and business analyses. He worked for this company as a director from 2004 to 2021. In 2005, he was selected NEDO Industry Fellowship Program and worked for Mie Technical License Organization to support research collaborations and start-up companies. In 2006, he joined Mie University Graduate School of Engineering as an assistant professor. He received a Ph.D. in Medical Science from Mie University in 2009. He was a visiting researcher at Cincinnati Children's Hospital Medical Center in 2012. He worked as an associate professor from 2017 to 2023 and is now a full professor at Mie University. From 2018 to 2013, he worked for Mie University Carrer Support Center as an Assistant to the president. From 2017 to 2021, he also worked as a visiting associate professor for Suzuka University of Medical Science, Japan. He is now a full professor at Mie University Graduate School of Engineering. His current research topics are medical informatics, medical imaging, image recognition, welfare information systems, ergonomics, evolutionary computations, and their application. He is a member of IEEE (and SMC Society), HIMSS, and some Japanese academic societies.



SATOSHI SHIRAE is currently a MS student at the Graduate School of Engineering, Mie University, Japan. He was a visiting researcher at Cincinnati Children's Hospital Medical Center in 2024. His research interests include deep learning, computational pathology and multimodal fusion.



SHYAM SUNDAR DEBSARKAR is currently a PhD student at the Prasath Lab, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, and at the Department of Computer Science, University of Cincinnati, USA. His research focuses on AI applications in Medical Imaging, Deep Learning, Large Language Models (LLMs), Vision-Language Models (VLMs), and Machine Learning. He completed his MS in Informatics at the Technical University of Munich (TUM), where he developed a strong foundation in Software Development and AI applications. Prior to starting his Ph.D., he has accumulated over 10 years of professional experience in software development, collaborating with diverse companies and clients across India and Europe. His current research interests are in exploring novel AI-driven methods to enhance diagnostic accuracy in healthcare, through the use of deep learning models. His goals are to drive advancements in AI technologies that improve patient outcomes and streamline clinical workflows.



BRUCE ARONOW is a John J. Hutton Chair Professor at the Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center (CCHMC). He also serves as co-director of the Computational Medicine Center at CCHMC. Since joining Cincinnati Children's in 1986, he built tools such as ToppGene that is widely used in bioinformatics, and his lab developed approaches that allow researchers to improve their understanding of (and find new treatments for) various medical conditions. His lab analyzed many different types of data, developed algorithms, and built websites and databases. These tools allow researchers from varying disciplines and backgrounds to analyze genetic and genomic data — either their own or that gathered from published sources — to model and perform new research about normal development and disease. His current research interests include leveraging multimodal AI models for unifying imaging, genomics, and clinical data for various biomedical informatics challenges.



SURYA PRASATH is a mathematician with two decades of experience in the research areas of image processing, computer vision, machine learning, deep learning and AI. He received his BSc Mathematics from the University of Madras, India in 2002 with distinction. He received his MSc, and PhD Mathematics from the Indian Institute of Technology Madras, India in 2004 and 2009. He was a Postdoctoral Fellow at the Department of Mathematics, University of Coimbra, Portugal from 2010 to 2011. From 2012 to 2015 he was with the Computational Imaging and VisAnalysis (CIVA) Lab at the Department of Computer Science, University of Missouri, USA as a Postdoctoral Fellow, and from 2016 to 2017 as an Assistant Professor. He is currently an Assistant Professor in the Division of Biomedical Informatics at the Cincinnati Children's Hospital Medical Center, and Departments of Biomedical Informatics, Pediatrics, Computer Science, at the University of Cincinnati from 2018. He had summer fellowships/visits at Kitware Inc. NY, USA, The Fields Institute, Canada, IPAM at the University of California Los Angeles (UCLA), USA, and Duke University, USA. His main research interests include applied mathematics, data science, image processing, computer vision, and machine learning with applications in various biomedical informatics domains. His current research focuses are in biomedical imaging informatics, and bioinformatics with AI/ML/DL techniques. His lab has over 100 publications in top-tier journals, conferences, such as Nature, Nature Communications, PLOS Computational Biology, Pediatric Research, IEEE Transactions on Image Processing, and more. He has received over \$2M research fundings from NIH, NSF, CCRF, CFF, etc. He is committed to promoting diversity in STEMM and has mentored several high-school, undergraduate, and graduate students in the last 15 years.

• • •